

Open Research Online

The Open University's repository of research publications and other research outputs

Library resources, student success and the distance-learning university

Journal Item

How to cite:

Nurse, Richard; Baker, Kirsty and Gambles, Anne (2018). Library resources, student success and the distance-learning university. *Information and Learning Science*, 119(1/2) pp. 77–86.

For guidance on citations see [FAQs](#).

© 2017 Emerald



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1108/ILS-03-2017-0022>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Library resources, student success and the distance-learning university

Richard Nurse, Kirsty Baker, Anne Gambles

Open University Library Services

Abstract

Purpose

Research at the Open University Library Services in the United Kingdom has been investigating the relationship between access to online library resources and student success to help to understand whether there is a similar relationship at a distance-learning university to that found in other institutions.

Design/methodology

The study analysed online library resource data from access logs from the EZproxy and OpenAthens systems. A data set of 1.7 million online resource accesses was combined with student success data for around 90,000 undergraduate students and a series of analyses undertaken.

Findings

The study found a pattern where students who are more successful are accessing more library resources. A Chi-square test indicated a statistically significant association between library resource accesses and module result, while an Anova test suggests a medium sized effect. The study also found that 152 (76%) of 199 modules had a small, medium or large positive correlation between student success, measured by the overall assessment score, and online library resource accesses.

Originality/value

This study builds on evidence that there is a relationship between library use and student success by showing that this relationship extends to the setting of a non-traditional, innovative library service supporting part-time distance learners.

Introduction

For many years the library of the Open University (OU), a United Kingdom-based distance learning higher education institution has collected statistics of access to online resources as a means of measuring use of subscription resources. The potential of these data to provide a greater depth of insight into the relationship between library online resource accesses and student success has become apparent through innovative work in libraries worldwide.

Studies at other academic libraries (Stone and Ramsden, 2013), (Cox and Jantti, 2012) have consistently found correlations between library use and student success. These studies led the library to ask if this pattern was also to be found at a distance learning institution where the student experience of 'the library' was an online presence rather than a physical building on a campus.

The OU has been developing capabilities around Learning Analytics for a number of years (Tynan and Mayles, 2014). This work included designing an ethical policy on the use of student data [i] and a framework around the use of learning analytics to guide interventions (Rienties *et al.*, 2016). Student engagement with library services is seen as part of the overall picture of learner engagement within these models.

Background

The Open University is a large distance learning institution, with over 170,000 students [ii]. Students are mainly part-time, and have a higher average age than at traditional 'campus-based' institutions. Students rarely visit the main university campus in Milton Keynes. The OU model of distance learning, called 'supported open learning', provides students with the materials they need to study, using a comprehensive mix of Virtual Learning Environment, printed and audio-visual course materials. Support is provided by a network of over 4,000 tutors. Students are able to communicate with their peers at tutorials, day schools, study groups, and via online channels.

The OU Library Service for students is delivered online, and provides access to an extensive collection of online resources as well as skills development and training materials. There is no 'physical' library service for students so library staff work intensively to embed the use of online library resources and digital and information literacy skills within OU modules (Reedy and Goodfellow, 2014).

Literature review

Research by academic libraries in the UK, the US and Australia has provided evidence that there is a positive relationship between different dimensions of library use and academic achievement.

The Library Impact Data Project based at the University of Huddersfield looked for evidence of a significant correlation between library activity data and student attainment across eight UK universities. Phase I of the project (Stone and Ramsden, 2013) investigated the final degree award against a selection of measures of library usage. A statistically significant relationship was identified across all eight universities between attainment and two library usage measures: online resource use and book borrowing. The project has undertaken further work including Phase 3 (Stone *et al.*, 2015), focusing on key findings from previous work to increase usage of library resources among low use groups.

Hong Kong Baptist University (Wong and Webb, 2010) explored the relationship between cumulative Grade Point Average (GPA) and library material usage (book and audio visual loans) for students who graduated within a 3 year period (8,701 cases). They divided the population into sample groups based on study major and level of study, and found that of 48 valid sample groups, 31 groups were statistically proven to have a positive relationship between library material use and GPA. Relationship strength ranged from small to medium. In addition, they noticed that for disciplines where they had identified weaker correlations between GPA and book/AV loans, their students had used more online resources.

The University of Wollongong appear to be unique in creating an ongoing system that joins library usage data to student systems data. Usage data comprised loans data and online resource usage data. Initial findings were that there was no direct correlation between library resource usage and academic performance. However, further investigations were undertaken looking at the central tendency of academic performance for a given frequency of usage, and comparing that to the average for other frequencies of usage. This

investigation revealed strong evidence that the library contributes to academic success (Cox and Jantti, 2012).

The University of Minnesota (Soria *et al.*, 2014) used GPA to investigate the relationship between first year undergraduates' use of the library and their achievement. The study looked at 10 different measures of library usage, and found that four aspects of library use were consistently and positively associated with higher GPA: database logins, book loans, online journal logins, and workstation logins (Soria *et al.*, 2014).

A further study (Soria *et al.*, 2017) acknowledged that to date studies have focussed on a limited set of academic outcomes, often GPA. This study was designed to investigate the association between library use and a wider set of academic outcomes: academic engagement, engagement in scholarly activities, and academic skills development. Results suggest that students using online library services had better academic engagement and higher grade point averages.

These studies and others have consistently demonstrated a relationship between aspects of library use and academic achievement. Several of these studies have examined use of online resources as a measure of library use and have demonstrated a positive correlation with academic success. However, these studies are from campus based institutions where students make use of a mixture of physical and online library resources and services. Our research aimed to investigate whether or not there was a similar relationship at a large distance learning institution where the library service is online.

Library data

Data about student access to online resources at the OU is derived from two different systems – OCLC EZproxy and Eduserve's OpenAthens online resource sign on system. EZproxy, a proxy server application, is used by many libraries as a means of managing 'off-campus' access to library subscription resources. As students are 'off-campus' at the OU, EZproxy provides a convenient access mechanism for the majority of online resource accesses. OpenAthens is used for a small number of online resources where IP authentication is not supported. As some courses make use of online resources that are accessed through OpenAthens, data from both systems has been combined by the project.

Both systems track student access but provide slightly different data outputs. OpenAthens provides a monthly summary report identifying the user, resource and a count of accesses per user/per resource. EZproxy provides a daily log file recording each transaction. The EZproxy Starting Point URL (SPU) log file was used as the basis for the analysis. This log file records a single transaction for each resource accessed by an individual user in a session. EZproxy does provide a complete log file of all transactions but this isn't routinely processed at the OU owing to its size. Unlike the analysis undertaken at Wollongong (Cox and Jantti, 2012) which made use of the larger log file to calculate an estimate of the time students spent accessing library resources, the OU analysis counts the number of accesses to resources.

Both the EZproxy and OpenAthens data were processed into a MySQL database and some additional data added including the module(s) being studied by the student at the date of the resource access and an additional student identifier to allow the data to be matched with student success data. Some work in Excel was undertaken to clean and prepare the data and join the data sources together.

Library data and student success data

Student results data from the institutional data warehouse provided several measures of student success. An Overall Continuous Assessment Score (OCAS) made up from the assessment scores throughout the module, an Overall Examination Score (OES), and a Result – variations of Fail, Pass, Distinction but also Withdrawn and Deferred. Broadly at the Open University 40% is required to gain a pass and 85% to gain a distinction in a module [iii].

The online library resource access count was added to the results data. Students without any recorded library resource accesses were marked as zero library resource accesses. A further piece of processing was undertaken at this stage to count the number of modules that students had started at the same time. This count was used to divide the resource access count (the rationale behind this approach is discussed in the challenges section below).

To simplify the presentation of the data several categories of results were merged together as some modules distinguish between different categories and others do not. Several fail categories were merged together as fail and a similar approach was taken with several variants of pass. This left a simpler fail, pass and distinction breakdown.

The project undertook an analysis of data from undergraduate students starting modules in autumn 2015. The data includes students who have studied previous modules at the OU and are starting their next module and those for whom this was their first experience of studying with the OU. A series of selections were made to provide a dataset for analysis. Only students studying modules at levels one, two and three were selected. Students without an Overall Continuous Assessment Score (including students who withdrew) and an Overall Examination Score, were removed from the dataset. Only students with a fail, pass or distinction result were included (therefore excluding students who deferred or had a pending result). Although this approach does exclude modules that do not have an examination as part of their assessment it does mean that the research was able to compare patterns of library resource access against both scores. Outlier values (as discussed below in the Statistical analysis findings section) were removed from the dataset. This left a dataset of 86,954 students.

Challenges

A particular challenge at the OU is the nature of the student journey. At the OU, undergraduate students study their degree through a number of separate 30 or 60 point modules and need to assemble 360 points from a mix of compulsory and optional modules

at levels one, two and three. The student journey is not a fixed path and students may be studying more than one module at a time, have overlapping modules or have study breaks.

Because students will often be studying more than one module at a time it is not feasible to relate a resource access to a specific module. When the access data is extracted from the library data store each transaction is included against each of the modules being studied by that student. An adjustment was made to divide the access count by the number of modules being studied by that student. The approach was verified with a sample set to check the validity. The adjusted access count was used throughout the study.

Findings

There are clearly a wide variety of factors that influence academic success. Correlation between library resource use and student marks isn't causation, as the Library Impact Data Project noted (Stone and Ramsden, 2013). It doesn't necessarily follow that the resource has been read, or understood. It may also be that students have found other routes to accessing the library resources they want to use, e.g. via other institutional affiliations.

The data showed that there is a consistent pattern of an increase in the mean number of online resources being accessed per student the higher the level of undergraduate study. Students studying level one modules averaged 7.5 online resource accesses, for level two modules the average was 18.7, and for level three modules 41.1. This means that there is a pattern of mean online resource accesses being nearly three times higher between level one and two and a little over two times between level two and level three.

The next step was to look at the breakdown by result and level of study. The data analysis uncovered a pattern that students who failed their module accessed an average of around 10 online resources, those who passed around 19 online resources and those gaining a distinction an average of 28 online resources. The pattern for the different levels is shown in Figure 1.

Figure 1

When a comparison was undertaken into the variations between pass/fail and distinction/pass for each of the levels it showed a variation between pass and fail at both levels one and two of just over twice the number of resource accesses. There was a lower variation between distinction and pass at levels two and three i.e. students at level two gaining a distinction are only accessing 1.3 times the number of online resources accessed by students who pass.

Further work allowed the data to be broken down by individual modules and faculties and quickly identified wide variations between different modules, providing a useful tool for Academic Liaison Librarians in conversations with academics. Once this pattern became apparent it was clear that some detailed statistical analysis would be needed to identify whether there was any statistical significance or whether it was just an interesting occurrence.

Statistical analysis findings

Discussions with colleagues and a review of the other approaches, notably the work by the University of Wollongong on the Library Cube (Cox and Jantti, 2012), identified several methods that could be used for the statistical analysis of the data.

One of the main features of the dataset is that the online resource access counts are not normally distributed. There are a large number of values of zero and a small number with high resource access counts. Analysis within SPSS showed it to have a positive skewed distribution as shown in Table 1.

Table 1

As there were a number of high value outliers some investigation was undertaken to test whether they were valid values or not. Using Wilcox’s formula (Wilcox, 2010) 9,920 values (11.4%) of over 47.5 were identified as being high outliers. However an investigation of these values indicated that many were from students who had completed and passed their modules, implying that they were valid values. After looking at the frequency of the access count values it was decided to remove values from the level where the majority of values were of a single frequency. This removed the outliers with scores of 600 or above (85 values i.e. around 0.1% of the total) from the analysis leaving a dataset of 86,954 student results, representing some 1.7 million online resource accesses.

As the percentage score was not available for the overall result a statistical analysis was undertaken using the OCAS and OES values, both presented as percentage scores. A scatterplot was created for the OCAS against the library online resource access count (Figure 2). This suggests that there is some relationship between the two values as there does seem to be a rising floor pattern, very similar to that seen in the Wollongong research (Cox and Jantti, 2012). The scatterplot for Overall Examination Score (OES) against library online resource accesses showed a similar pattern.

Figure 2

In the next step Spearman’s rho correlation calculations were carried out in SPSS for the online resource access count against the OCAS and OES values. These provided a Spearman’s rho result of $r_s(86,954) = .082$, $p < .0005$, significant at the 0.01 level (2-tailed) for the OCAS value and $r_s(86,954) = .071$, $p < .0005$, significant at the 0.01 level (2-tailed) for the OES value. Breaking this down for the level one, two and three modules found values ranging between .076 and .137 for the OCAS value and .095 and .130 for the OES value.

Data for 199 individual modules was analysed in the same way and the results categorised according to Cohen (Cohen, 1992). The results showing the number of modules by Cohen category are shown in Table 2.

Table 2

The correlation r_s values for the OCAS and OES values where there was a significant correlation for both (140 modules with $p < 0.05$ for both correlations) were plotted on a scatterplot (Figure 3). The correlation analysis showed an interesting pattern suggesting a clear relationship between the number of resources being accessed and the module result.

Figure 3

In the next test, box plots were used to identify all high outliers. All resource access counts above 267 were removed as outliers. The resource access counts were split into zero (0), very low (0.01 -2.00), low (2.01 – 6.00), medium (6.01 – 14.67), high (14.68 – 36.00) and very high (36.01 and above) bands, a total of 86,375 valid cases. A Chi-Square test of independence was conducted between resource access band and module result. All expected cell frequencies were greater than five. There was a statistically significant association between resource access band and module result, $\chi^2(10) = 796.216$, $p < .0005$.

As a final stage in this phase of the project, resource access counts of zero were removed, giving a dataset of 61,971 students. After transforming the resource access counts to a logarithmic value, an Anova test was undertaken comparing the resource access count against module result. This gave a Sum of Squares value of 976.571 (between groups) and 97903.464 (within groups), an F value of 309.060 at a significance level of .001 with an Eta squared value of .010, suggesting a medium effect size between the groups.

Discussion

The initial analysis showing rising amounts of resource accesses as students study at higher levels seems to fit with a pattern at the OU where students are expected to be using library resources much more at levels two and three. The analysis proved to be valuable as an advocacy tool in conversations with senior stakeholders about library data and was included as part of a new annual report provided to Schools. The results shown in Figure 1 suggest that, at the very least, there is an interesting relationship between student success and library engagement. Students who are getting better results are accessing more online library resources and students who are failing are accessing fewer online library resources and this pattern goes across the different levels of study.

This is also suggested by Figure 2 showing a rising pattern for the OCAS value against library resource accesses. This pattern seems to be quite similar to that seen by the research undertaken by Wollongong (Cox and Jantti, 2012).

The analyses undertaken for each module (shown in Table 2) indicate that around three-quarters of modules have some form of correlation between the number of library resources accessed and the continuous assessment score. This large sample supports the view that students gaining higher marks were accessing more library resources. The later statistical tests also reinforce the view that there is the same sort of statistically significant relationship between online library resource accesses and student results at the OU as found by Huddersfield (Stone and Ramsden, 2013).

As previously noted OU students are typically older than traditional undergraduate students (although recently the number of younger students studying with the OU has

increased). There is some suggestion (Wilkinson, 2004) that older students may achieve better results and also that older age might be a predictor of success in web-based courses (Vella *et al.*, 2016). In the library sector, research by the Library Impact Data Project found that mature students (i.e. students aged 21 and over) tend to have higher online library resource usage (Stone and Collins, 2013). However work at Wollongong (Cox and Jantti, 2012) suggests that older students do not benefit from increased library use. Further analysis of the dataset broken down by age groups would be a potential future piece of work to help understand the extent to which age is a factor at the OU. Other demographic factors such as disability, ethnicity and previous educational achievement would be potential areas for future investigation as a follow on piece of work.

Conclusions

At the OU there is a clear pattern of an increase in the average number of library resources being accessed by students as they rise through the undergraduate levels of study. Students who attain higher grades are accessing a larger number of library resources across all levels.

Although level 1 students are accessing fewer library resources it is still the case that students who pass at level 1 access more library resources than those who fail and students gaining a distinction access more resources than students who pass. Other studies (Soria, 2014) have suggested that early library engagement is beneficial. This research seems to support the view that an approach that encourages students to use library resources earlier may be beneficial to student success.

There is considerable variation between modules. This comes through strongly in that although there is no across the board correlation between library online resource accesses and student success at the OU, there are 152 modules (out of 199) with a small, medium or large correlation between the continuous assessment score and library resource accesses and 150 modules with a correlation with the overall examination score.

The findings are clearly applicable to other similar distance-learning and online learning institutions. The correlation between library resource accesses and student success suggests that the OU is similar to other higher education institutions that have campus-based student populations. That the OU shows a similar pattern to that identified at Huddersfield, Minnesota and others is clearly significant as it suggests that students, who only have access to an online library, experience similar benefits to students who are studying at campuses with physical libraries. As one of the big challenges for university libraries is to understand the implications of a shift from print to digital, there is a strong suggestion here that a primarily digital environment does not disadvantage students. That may be reassuring to other institutions contemplating the implications of a digital library world.

Given the limitations of correlational measures further statistical tests should be able to shed more light on the nature of the relationship between resource accesses and student success. This could include undertaking further tests for the different levels and modules, including Anova, Chi-square tests and potentially undertaking ordinal regression analyses.

- i. <http://www.open.ac.uk/students/charter/essential-documents/ethical-use-student-data-learning-analytics-policy> (accessed 30 January 2017).
- ii. <http://www.open.ac.uk/about/main/strategy/facts-and-figures> (accessed 30 January 2017).
- iii. <http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/assessment-handbook-after290912.pdf> (accessed 30 January 2017).

References

Cohen, J. (1992), "A power primer", *Psychological Bulletin*, Vol. 112, No. 1, pp. 155-159, available at: <http://www2.psych.ubc.ca/~schaller/528Readings/Cohen1992.pdf> (accessed 2 February 2017).

Cox, B. L. and Jantti, M. (2012), "Capturing business intelligence required for targeted marketing, demonstrating value, and driving process improvement", *Library & Information Science Research*, Vol. 34, No. 4, pp. 308–316, (accessed 2 February 2017).

Reedy, K. and Goodfellow, R. (2014), "'You've been frameworked': evaluating an approach to digital and information literacy at the Open University", *Journal of Learning Development in Higher Education*, Vol. 7, available at: <http://oro.open.ac.uk/41547/> (accessed 2 February 2017).

Rienties, B., Boroowa, A., Cross, S., Kubiak, C., Mayles, K., and Murphy, S. (2016), "Analytics4Action evaluation framework: a review of evidence-based learning analytics interventions at the Open University UK", *Journal of Interactive Media in Education*, Vol. 2016(1), available at: <http://www-jime.open.ac.uk/articles/10.5334/jime.394/> (accessed 2 February 2017).

Soria, K. M., Fransen, J. and Nackerud, S. (2014), "Stacks, serials, search engines, and students' success: first-year undergraduate students' library use, academic achievement, and retention", *The Journal of Academic Librarianship*, Vol. 40, No. 1, pp. 84–91, (accessed 2 February 2017).

Soria, K. M., Fransen, J. and Nackerud, S. (2017), "Beyond books: the extended academic benefits of library use for first-year college students", *College & Research Libraries*, Vol. 78, No. 1, available at: <http://crl.acrl.org/index.php/crl/article/view/16564> (accessed 13 June 2017).

Stone, G. and Collins, E., (2013), "Library usage and demographic characteristics of undergraduate students in a UK university", *Performance Measurement and Metrics*, Vol. 14, No. 1, pp. 25–35, (accessed 14 June 2017).

Stone, G. and Ramsden, B. (2013), "Library impact data project: looking for the link between library usage and student attainment", *College & Research Libraries*, Vol. 74, No. 6, pp. 546–559, available at: <http://crl.acrl.org/content/74/6/546> (accessed 2 February 2017).

Stone, G., Sharman, A., Dunn, P. and Woods, L. (2015), "Increasing the impact: building on the library impact data project", *The Journal of Academic Librarianship*, Vol. 41, No. 4, pp. 517–520, available at: <http://eprints.hud.ac.uk/24917/> (accessed 2 February 2017).

Tynan, B., and Mayles, K. (2014), *Transforming assessment webinar series: learning analytics*, available at: http://transformingassessment.com/sites/default/files/files/TA_webinar_25_june_2014_slides_learning_analytics.pdf (accessed 30 January 2017).

Vella, E. J., Turesky, E. F., and Hebert, J. (2016), "Predictors of academic success in web-based courses: age, GPA, and instruction mode", *Quality Assurance in Education*, Vol. 24 No. 4, pp.586–600, (accessed 14 June 2017).

Wilcox, R. R. (2010), *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*, 2nd edition, Springer, New York.

Wilkinson, T.J., Wells, J.E. and Bushnell, J.A. (2004), "Are differences between graduates and undergraduates in a medical course due to age or prior degree?", *Medical Education*, Vol. 38, pp.1141–1146, (accessed 14 June 2017).

Wong, S. H. R. and Webb, T. D. (2010), "Uncovering meaningful correlation between student academic performance and library material usage", *College & Research Libraries*, Vol. 72, No. 4, pp. 361–370, available at: <http://crl.acrl.org/content/72/4/361> (accessed 2 February 2017).

Figure 1

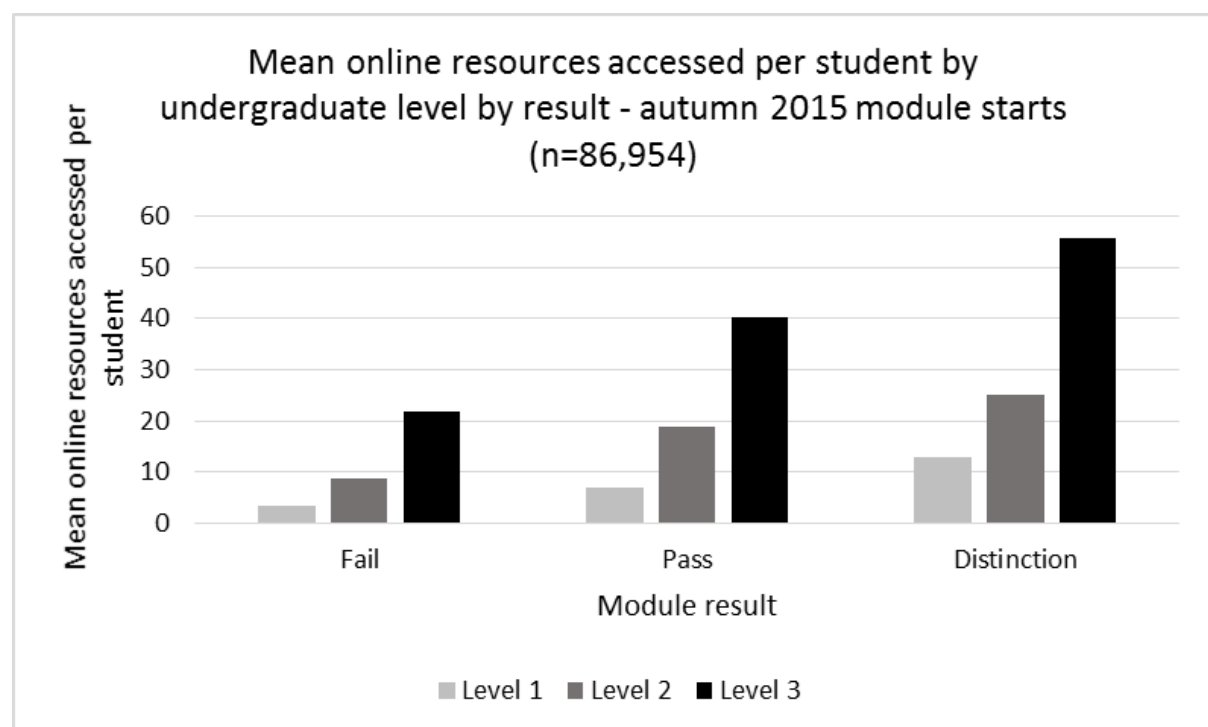


Table 1

Descriptives				Statistic	Std. Error
Count	Mean			19.9243	.15143
	95% Confidence Interval for Mean	Lower Bound		19.6275	
		Upper Bound		20.2211	
	5% Trimmed Mean			12.4208	
	Median			4.0000	
	Variance			1993.931	
	Std. Deviation			44.65346	
	Minimum			.00	
	Maximum			590.00	
	Range			590.00	
	Interquartile Range			19.00	
	Skewness			5.050	.008
	Kurtosis			35.422	.017

Figure 2

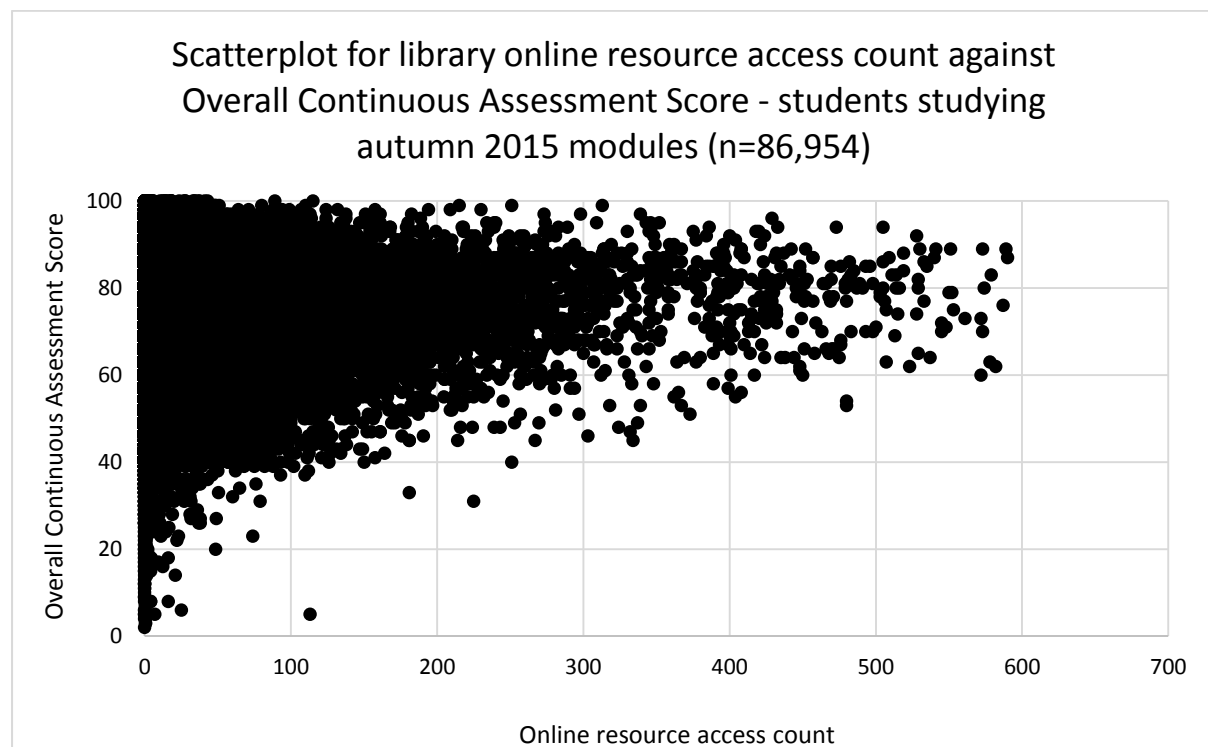


Table 2

Number of modules with Spearman's rho correlation between Overall Continuous Assessment Score and online resource accesses using Cohen categorisation.					
	None $r_s. < .1$	Small $r_s. \geq .1 \text{ \& } < .3$	Medium $r_s. \geq .3 \text{ \& } < .5$	Large $r_s. \geq .5$	Not significant $p > 0.05$
Level 1 (n=51)	2 (4%)	24 (47%)	13 (25%)	0 (0%)	12 (24%)
Level 2 (n=81)	0 (0%)	30 (37%)	34 (42%)	1 (1%)	16 (20%)
Level 3 (n=67)	0 (0%)	18 (27%)	32 (48%)	0 (0%)	17 (25%)

Number of modules with Spearman's rho correlation between Overall Examination Score and online resource accesses using Cohen categorisation.					
	None $r_s. < .1$	Small $r_s. \geq .1 \text{ \& } < .3$	Medium $r_s. \geq .3 \text{ \& } < .5$	Large $r_s. \geq .5$	Not significant $p > 0.05$
Level 1 (n=51)	2 (4%)	30 (59%)	5 (10%)	0 (0%)	14 (27%)
Level 2 (n=81)	3 (4%)	36 (44%)	22 (27%)	0 (0%)	20 (25%)
Level 3 (n=67)	0 (0%)	33 (49%)	24 (36%)	0 (0%)	10 (15%)

Figure 3

